# Hierarchical Texture Motifs

Shawn Newsam

School of Engineering
University of California
Merced, CA 94344, U.S.A.

## ABSTRACT

A fundamental challenge in analyzing spatial patterns in images is the notion of scale. Texture based analysis typically characterizes spatial patterns only at the pixel level. Such small scale analysis usually fails to capture spatial patterns that occur over larger scales. This paper presents a novel solution, termed **hierarchical texture motifs**, to this texture-of-textures problem. Starting at the pixel level, spatial patterns are characterized using parametric statistical models and unsupervised learning. Higher levels in the hierarchy use the same analysis to characterize the motifs learned at the lower levels. This multi-level analysis is shown to outperform single-level analysis in classifying a standard set of image textures.

**Keywords:** Texture features, texture classification

## 1. INTRODUCTION

This work presents a novel approach to the problem of scale in the analysis of spatial patterns in images. Image texture can be thought of spatial patterns at the pixel level. These are small scale patterns. Spatial patterns that occur at larger scales are difficult to characterize using pixel based approaches. Consider, for example, texture D017 in figure 5. Pixel level analysis might be able to characterize the texture of the two different directions of weave but is unlikely to be able to characterize the spatial relationship of the two weaves (for example, as parallel stripes). Some kind of texture-of-textures approach is needed to simultaneously capture the spatial patterns that occur at different scales.

We here present a hierarchical approach to this problem in which different scales of spatial patterns are analyzed at different levels in the hierarchy. A salient feature of this approach is that the same analysis is performed at each level or scale. At the lowest level, texture analysis is performed at the pixel scale. This results in a symbolic image whose pixel labels correspond to texture classes. The next level in the hierarchical analysis simply considers the integral labels as gray scale values and repeats the same analysis, now at a larger scale. We apply this approach to a set of 25 commonly used texture images and show that additional levels result in significant improvement in classification accuracy.

The rest of the paper is organized as follows. Section 2 describes previous work; section 3 describes the texture features; section 4 describes texture motifs; section 4 describes the hierarchical extension to texture motifs; section 5 describes the experimental results; and section 6 concludes with a discussion.

## 2. RELATED WORK

Simultaneously characterizing both the pixel level spatial patterns, or pixel texture, and the spatial relationships of the textures themselves has been acknowledged as a challenging problem for some time. One set of techniques solve this problem by modelling the spatial relationships of the textures using hidden Markov models or Markov random fields so that the texture label assigned to a pixel depends on the labels assigned to neighboring pixels. This field effect can either be incorporated directly into the learning of the texture classes as constraints in a neural network classifier,[1] or as state-conditioned feature distributions,[2] or as a postprocessing step.[3] While these approaches for considering additional levels of spatial interaction have shown to be effective, they are significantly more complex than the hierarchical approach presented here in which the same model is applied

Further author information: S.N.: E-mail: snewsam@ucmerced.edu

Image Processing: Algorithms and Systems V, edited by Jaakko T. Astola, Karen O. Egiazarian, Edward R. Dougherty
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6497, 649708, © 2007 SPIE-IS&T · 0277-786X/07/$18

recursively. Another approach, motivated by Julesz notion of textons,[4] uses a generative image model in which the spatial primitives in natural images are considered to be composed of simpler bases such as filter outputs.[5] However, this approach only provides a multi-level decomposition of the spatial primitives; it does not consider the spatial relationships of the primitives themselves and therefore does not truly address the texture-of-textures problem.

## 3. IMAGE TEXTURE FEATURES

The proposed approach assumes that texture feature vectors extracted from gray scale images using spatial filters tuned to combinations of $R$ orientations and $S$ scales are used to characterize the spatial patterns:

$$\begin{aligned} x^n &= [(x^n_{11}, x^n_{12}, \ldots, x^n_{1S}), (x^n_{21}, x^n_{22}, \ldots, x^n_{2S}), \\ &\quad \ldots, (x^n_{R1}, x^n_{R2}, \ldots, x^n_{RS})] \,. \end{aligned} \tag{1}$$

Here, element $x^n_{rs}$ represents the output at pixel location $n$ of the filter tuned to orientation $r$ and scale $s$. Parentheses are used to group the filter outputs by orientation for clarity. A typical filter bank consists of filters tuned at $180/R$ degree intervals.

## 4. TEXTURE MOTIFS

A texture motif is here defined as *a characteristic spatial pattern common to a class of objects*.[6] The pattern can occur at different locations and orientations within the objects. Examples of texture motifs include the rows of moored boats in aerial images of harbors and the rows of trees in aerial images of golf courses. Accurate characterizations of texture motifs can facilitate automated object recognition. However, developing the characterizations is a challenge largely due to the high-dimensionality of the texture feature spaces. In the proposed approach, the distribution of the feature vectors corresponding to a texture motif is characterized with a parametric statistical model whose parameters are estimated from unlabeled training samples using an unsupervised learning algorithm. The advantages of this approach are 1) preprocessing steps, such as segmentation are not required; 2) the learning phase is completely automated, only requiring unlabeled training examples; and 3) the model form is fixed so that only the parameter values differ from the model of one class of objects to another. Texture motifs have shown to be effective at characterizing geospatial objects in remote sensed imagery.[6] Techniques have also been developed to make them scale and orientation[7] invariant, even when the underlying feature vectors are not, which are often an important criteria when dealing with real imagery.

### 4.1. STATISTICAL MODELING VIA GMMS

The feature vectors corresponding to a texture motif are assumed to have a Gaussian distribution in the $RS$ dimensional texture feature space. The conditional probability of a feature vector $x$, given that it is generated by texture motif $j$, is thus

$$p\left(x|j\right) = \frac{1}{\left(2\pi\right)^{d/2}\left|\Sigma_j\right|^{1/2}}\,\mathrm{e}^{-\frac{1}{2}(x-\mu_j)^T\Sigma_j(x-\mu_j)} \tag{2}$$

where $d = RS$. The density for motif $j$ is completely specified by the parameters $(\mu_j, \Sigma_j)$, where $\mu_j$ is the mean vector and $\Sigma_j$ is the covariance matrix. The density of an ensemble of $J$ texture motifs is consequently modeled using a Gaussian Mixture Model (GMM), so that the unconditional probability of a feature vector $x$, with respect to an ensemble of motifs, can be computed as

$$p\left(x\right) = \sum_{j=1}^{J} P\left(j\right) p\left(x|j\right) \tag{3}$$

where $P(j)$ is the prior probability of motif $j$ and $p(x|j)$ is the conditional probability. The model for an ensemble of $J$ motifs is completely specified by the parameters

$$\Theta = \left\{\theta_j = \left(P\left(j\right), \mu_j, \Sigma_j\right) | j = 1 \ldots J\right\} \,. \tag{4}$$

## 4.2. UNSUPERVISED LEARNING VIA THE EM ALGORITHM

The Expectation Maximization (EM) algorithm[8] is a common technique for learning the parameters of a statistical model from an unlabeled training set. It is an iterative learning technique in which the observable data is augmented by the missing elements needed to estimate the values of the model parameters. The missing elements for GMMs are the mixture assignments of the feature vectors, $z \in J$. Estimating the parameter values $\{\theta_j | j = 1 \dots J\}$ from the training set is straightforward if these assignments are known (using maximum-likelihood, for example). The complete data for the GMM case is $y = (x, z)$, where $x$ is the observable feature vector and $z$ is the unknown mixture assignment.

Each iteration $m$ of the EM procedure has two steps. First, the current estimates of the parameters are used to compute the expected value of the log-likelihood of the complete training data given the observed data where the expectation is with respect to the unknown data. Second, the parameter estimates are updated to maximize this expectation. These two steps are repeated until a stopping criterion is met, often related to the rate of change of the likelihood.

For the GMM case, the current estimates at iteration $m$ are $\Theta^{(m)} = \left\{\theta_j^{(m)} | j = 1 \dots J\right\}$; the complete data are $Y = (X, Z) = \{(x^n, z^n) | n = 1 \dots N\}$; the observed data are $X = \{x^n | n = 1 \dots N\}$; and the unknown data are $Z = \{z^n | n = 1 \dots N\}$, where $N$ is the size of the training set. The expected value of the log-likelihood of the training data is commonly written as

$$Q\left(\Theta|\Theta^{(m)}\right) = E_Z\left[\log p\left(Y|\Theta\right)|X, \Theta^{(m)}\right] \ . \tag{5}$$

The maximization step involves updating the parameter estimates to the values of $\Theta$ that maximize $Q$. Fortunately, this optimization problem has an analytical solution for the GMM case, and the updated estimates for the model parameters are computed as

$$P^{(m+1)}\left(j\right) = \frac{1}{N}\sum_{n=1}^{N} P^{(m)}\left(j|x^n\right) , \tag{6}$$

$$\mu_j^{(m+1)} = \frac{\sum_{n=1}^{N} P^{(m)}\left(j|x^n\right) x^n}{\sum_{n=1}^{N} P^{(m)}\left(j|x^n\right)} , \tag{7}$$

and

$$\Sigma_j^{(m+1)} = \frac{\sum_{n=1}^{N} P^{(m)}\left(j|x^n\right)\left(x^n - \mu_j^{(m+1)}\right)\left(x^n - \mu_j^{(m+1)}\right)^T}{\sum_{n=1}^{N} P^{(m)}\left(j|x^n\right)} . \tag{8}$$

The k-means clustering algorithm is typically used to initialize the parameter values, $\Theta^{(0)}$.

A diagram of learning single-level motif models is shown in figure 1.

## 4.3. HIERARCHICAL TEXTURE MOTIFS

The previous section described how the texture motifs are learned in an unsupervised fashion. The major contribution of this paper is to extend this single-level analysis by labeling the motifs and *treating these labeled images as gray scale images and repeating the texture feature extraction and motif modeling process.*

The learned motif models, represented by the GMMs, are used to assign a motif label to each pixel in an image using a maximum a posteriori (MAP) classifier:

$$\text{motif} = \underset{1 \le j \le J}{\arg\max}\, P\left(j|x\right) \ . \tag{9}$$

This results in an image with the same dimensions but with pixels with numeric labels $j$ ($1 \le j \le J$ where $J$ is the number of mixtures in the GMM). This motif labeled image can be considered as a gray scale image whose pixel values are simply the numeric motif labels. The same steps described in the previous section can then be used to *learn the texture motifs of the motif labeled images*. In particular, the motif labeled images are filtered with the same spatial filters, and the motifs are learned using the same GMM/EM framework.

A diagram showing how hierarchical motif models are learned in presented in figure 2. The complete model in this case is the combination of the first-level and second-level models.

# 5. EXPERIMENTAL RESULTS

The single-level and hierarchical texture motif models are compared using a set of standard texture images.

## 5.1. THE DATASET

A set of 25 texture images from the Brodatz texture dataset[9] are used to compare the single-level and hierarchical texture motif models: D016, D017, D020, D021, D034, D047, D053, D055, D056, D057, D064, D065, D076, D077, D082, D083, D085, D095, D101, D102, D103, D014, D109, D110, and D111. Each image measures 512x512 pixels and has gray scale values that range from 0 to 255. Figure 5 shows these 25 texture images (the center 256x256 pixels are shown for detail).

## 5.2. THE TEXTURE FEATURES

Texture feature vectors are extracted using a bank of Gabor filters tuned to $R$=6 orientations and $S$=5 scales. An attractive mathematical property of Gabor functions is that they minimize the joint uncertainty in space and frequency.[10] They achieve the optimal tradeoff between localizing the analysis in the spatial and frequency domains. Using Gabor filters to analyze texture also appeals from a psycho-visual perspective. In particular, researchers have found that Gabor functions can be used to model the receptive fields of simple cells in the mammalian visual cortex.[11] Such a finding suggests that Gabor-like filtering takes place early on in the early human visual system. Gabor texture features were standardized by the MPEG-7 Multimedia Content Description Interface,[12] and have shown to be effective at characterizing texture in a variety of image types, such as remote sensed imagery[13][14],[15] or the output of physics simulations.[16] Note, however, that the proposed approach is general enough to used with a variety of texture features extracted through spatial filtering.

## 5.3. THE MODELS

GMMs with five mixture components in the single-level motif model. The second level in the hierarchical model also used GMMs with five components. The mixture components in the GMMs have diagonal covariance matrices. The GMM/EM framework was used to learn the model parameters for each texture using a random sample of 10,000 feature vectors from the images (the computational complexity of the EM algorithm prevents using more features). This training set represents approximately 4.5% of the features in the 512x512 pixel images after the borders have been masked to prevent edge effects from the spatial filters.

## 5.4. EVALUATING THE MODELS

The single-level and hierarchical motif models were evaluated by comparing how well they classified test images. In particular, given a test image, we computed the likelihood that it was generated by each of the 25 models. For each pixel in the test image, we used equation 3 to calculate the probability of the pixel with respect to the each of the models. For the hierarchical motif models, *the product of the probabilities of each level was used*. A pixel was considered to be classified correctly when this probability was maximal for the model trained on the same texture. The number of correctly classified pixels was summed for a test image. This count was used to compare the single-level and hierarchical motif models on a per-test image bases as well as averaged over all test images.

Figures 3 and 4 demonstrate how the single-level and hierarchical models are used to classify pixels.

### 5.5. RESULTS

The 25 texture images are shown in figure 5. A single-level motif model with 5 components was learned for each of these images using 10,000 randomly sampled feature vectors (less than 5% of the total feature vectors). These models were then used to assign motif labels using a MAP classifier. The results of this labeling are shown in figure 6. These motif labeled images were then considered as gray scale images and the feature extraction and model learning process was repeated. This resulted in the second level of the hierarchical models.

The single-level and hierarchical models were evaluated as follows. For each of the 25 texture images, the probability of each pixel was computed with respect to the 25 single-level and hierarchical models. The number of pixels whose probability was maximal for the correct model was separately summed for the single-level and the hierarchical models. Table 1 lists these counts along with what percentage of the whole image they represent. The average count and percentage over all 25 images is also shown. These results demonstrate that the proposed hierarchical models result in a 5% classification improvement over the entire texture dataset.

## 6. CONCLUSION

This paper presents a texture-of-textures approach to characterize spatial patterns at multiple scales. An attractive feature of the proposed hierarchical texture motifs is that the analysis performed at each level is the same–just the image scales and interpretations differ. A two-level model is shown to improve the pixel classification of standard texture images by an average of 5% over a single-level model.

There is plenty of future work to be done. First, we would like to investigate how to pick the optimal number of mixtures for the models at the different levels. Initial investigations into using minimum description length show promise. We would also like to investigate how best to label the pixels when moving up to the next level of analysis. This includes issues such as what should the ordering and range of the integral values be. Sub-sampling the motif labeled images might also facilitate the characterization spatial patterns at different scales. Finally, we would like to explore whether adding a third and possibly additional levels further improves the models.

## REFERENCES


1. P. P. Raghu, R. Poongodi, and B. Yegnanarayana, "Unsupervised texture classification using vector quantization and deterministic relaxation neural network," *IEEE Transactions in Image Processing* **6**(10), pp. 1376–1387, 1997.
2. S. Newsam, S. Bhagavathy, and B. S. Manjunath, "Modeling object classes in aerial images using hidden markov models," in *IEEE International Conference on Image Processing*, **1**, pp. 860–863, 2002.
3. S. Newsam, S. Bhagavathy, and B. S. Manjunath, "Object localization using texture motifs and markov random fields," in *IEEE International Conference on Image Processing*, **2**, pp. 1049–1052, 2003.
4. B. Julesz, "Textons, the elements of texture perception and their interaction," *Nature* **290**(5802), pp. 91–97, 1981.
5. S.-C. Zhu, C. en Guo, Y. Wang, and Z. Xu, "What are textons?," *International Journal of Computer Vision* **62**(1-2), pp. 121–143, 2005.
6. S. Bhagavathy, S. D. Newsam, and B. S. Manjunath, "Modeling object classes in aerial images using texture motifs," in *International Conference on Pattern Recognition*, **2**, pp. 981–984, 2002.
7. S. Newsam and B. S. Manjunath, "Normalized texture motifs and their application to statistical object modeling," in *IEEE International Conference on Computer Vision and Pattern Recognition: Workshop on Perceptual Organization in Computer Vision*, 2004.
8. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *J. of the Royal Statistical Society* **39**(1), pp. 1–38, 1977.
9. P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover, New York, U.S.A., 1966.
10. J. Daugman, "Complete discrete 2D Gabor transform by neural networks for image analysis and compression," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36**, pp. 1169–1179, 1988.
11. S. Marcelja, "Mathematical description of the responses of simple cortical cells," *Journal of the Optical Society of America* **70**(11), pp. 1297–1300, 1980.

| Image | One Level | Two Levels |
|---|---|---|
| **D016** | 220712 (99.9%) | 220494 (99.8%) |
| **D017** | 217645 (98.5%) | 212688 (96.3%) |
| **D020** | 218926 (99.1%) | 219818 (99.5%) |
| **D021** | 220893 (99.9%) | 220897 (99.9%) |
| **D034** | 220759 (99.9%) | 220247 (99.7%) |
| **D047** | 211897 (95.9%) | 212198 (96.1%) |
| **D053** | 220859 (99.9%) | 219907 (99.5%) |
| **D055** | 211028 (95.5%) | 209022 (94.6%) |
| **D056** | 207672 (94.0%) | 208451 (94.4%) |
| **D057** | 200276 (90.6%) | 206116 (93.3%) |
| **D064** | 173177 (78.4%) | 192142 (87.1%) |
| **D065** | 181559 (82.2%) | 200181 (90.6%) |
| **D076** | 216594 (98.1%) | 216119 (97.8%) |
| **D077** | 220812 (99.9%) | 219467 (99.4%) |
| **D082** | 212826 (97.2%) | 214679 (97.2%) |
| **D083** | 216925 (98.2%) | 212606 (96.2%) |
| **D085** | 217132 (98.3%) | 210631 (95.4%) |
| **D095** | 190410 (86.2%) | 206555 (93.5%) |
| **D101** | 143485 (65.0%) | 216096 (97.8%) |
| **D102** | 129277 (58.5%) | 217916 (98.6%) |
| **D103** | 130872 (59.2%) | 175889 (79.6%) |
| **D104** | 135000 (61.1%) | 159705 (72.3%) |
| **D109** | 104114 (47.1%) | 124750 (56.5%) |
| **D110** | 124894 (56.5%) | 143986 (65.2%) |
| **D111** | 204674 (92.7%) | 211905 (95.9%) |
| **Average** | 190096 (86.1%) | 202989 (91.2%) |

**Table 1.** The number of pixels (and percent of image) classified correctly for the single-level and hierarchical motif models.


12. B. S. Manjunath, P. Salembier, and T. Sikora, eds., *Introduction to MPEG7: Multimedia Content Description Interface*, John Wiley & Sons, 2002.
13. S. Newsam, L. Wang, S. Bhagavathy, and B. S. Manjunath, "Using texture to analyze and manage large collections of remote sensed image and video data," *Journal of Applied Optics: Information Processing* **43**(2), pp. 210–217, 2004.
14. S. Newsam and C. Kamath, "Retrieval using texture features in high resolution multi-spectral satellite imagery," in *SPIE Defense and Security Symposium, Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI*, 2004.
15. S. Newsam, L. Wang, S. Bhagavathy, and B. S. Manjunath, "Using texture to annotate remote sensed datasets," in *International Symposium on Image and Signal Processing and Analysis, Special Session on Texture Analysis and Synthesis*, **1**, pp. 72–77, 2003.
16. S. Newsam and C. Kamath, "Comparing shape and texture features for pattern recognition in simulation data," in *SPIE International Symposium on Electronic Imaging, Image Processing: Algorithms and Systems IV*, 2005.

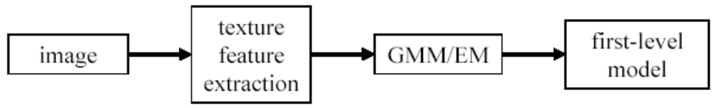


**Figure 1.** Learning single-level motif models.

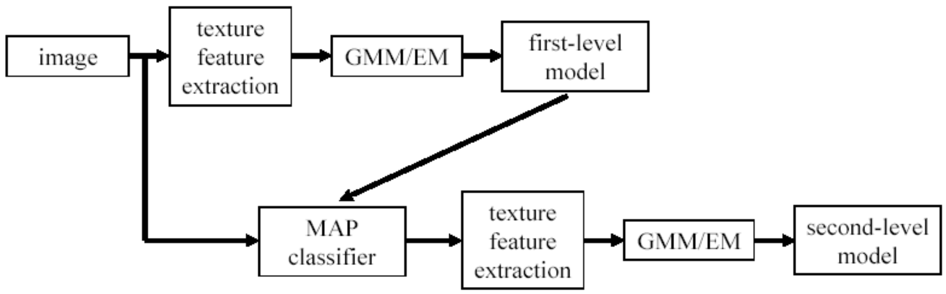


**Figure 2.** Learning multiple-level motif models.

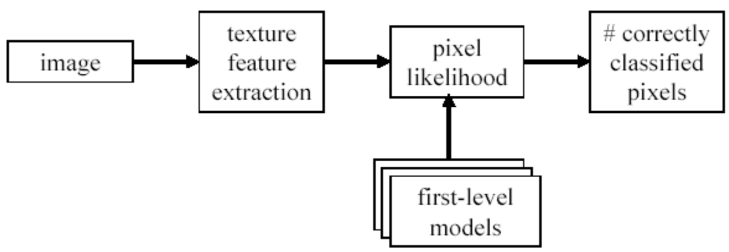


**Figure 3.** Classifying pixels using single-level motif models.

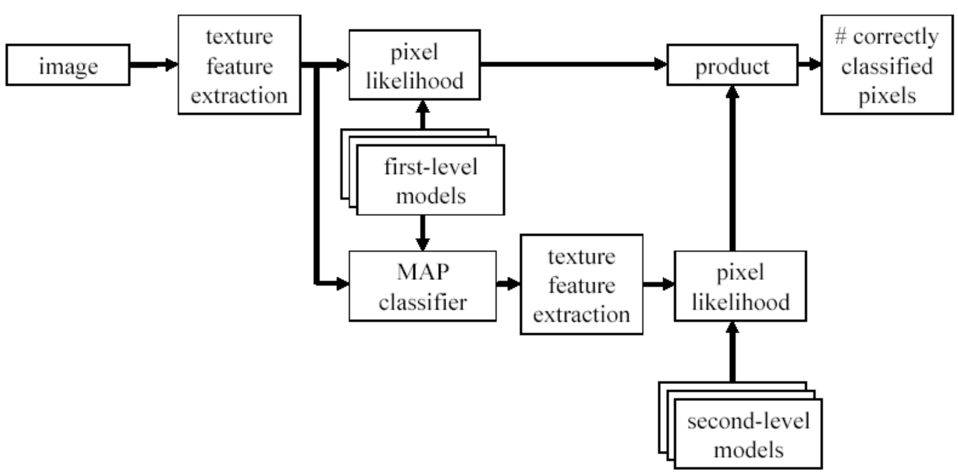


**Figure 4.** Classifying pixels using multiple-level motif models.

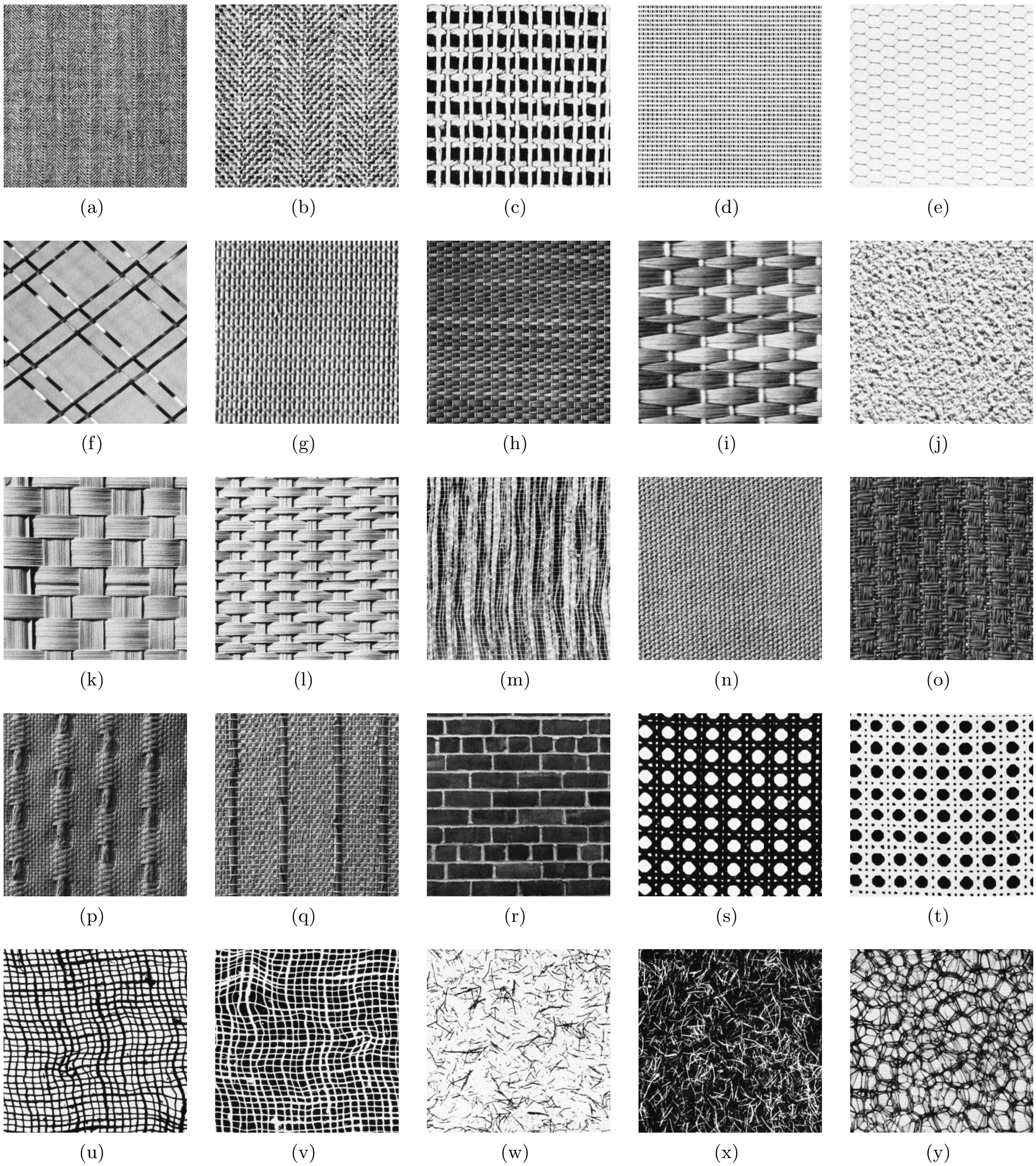


**Figure 5.** The 25 images from the Brodatz texture dataset[9] used to evaluate single and hierarchical texture motif models: (a) D016, (b) D017, (c) D020, (d) D021, (e) D034, (f) D047, (g) D053, (h) D055, (i) D056, (j) D057, (k) D064, (l) D065, (m) D076, (n) D077, (o) D082, (p) D083, (q) D085, (r) D095, (s) D101, (t) D102, (u) D103, (v) D014, (w) D109, (x) D110, and (y) D111. Only the center 256x256 pixels are shown for detail.

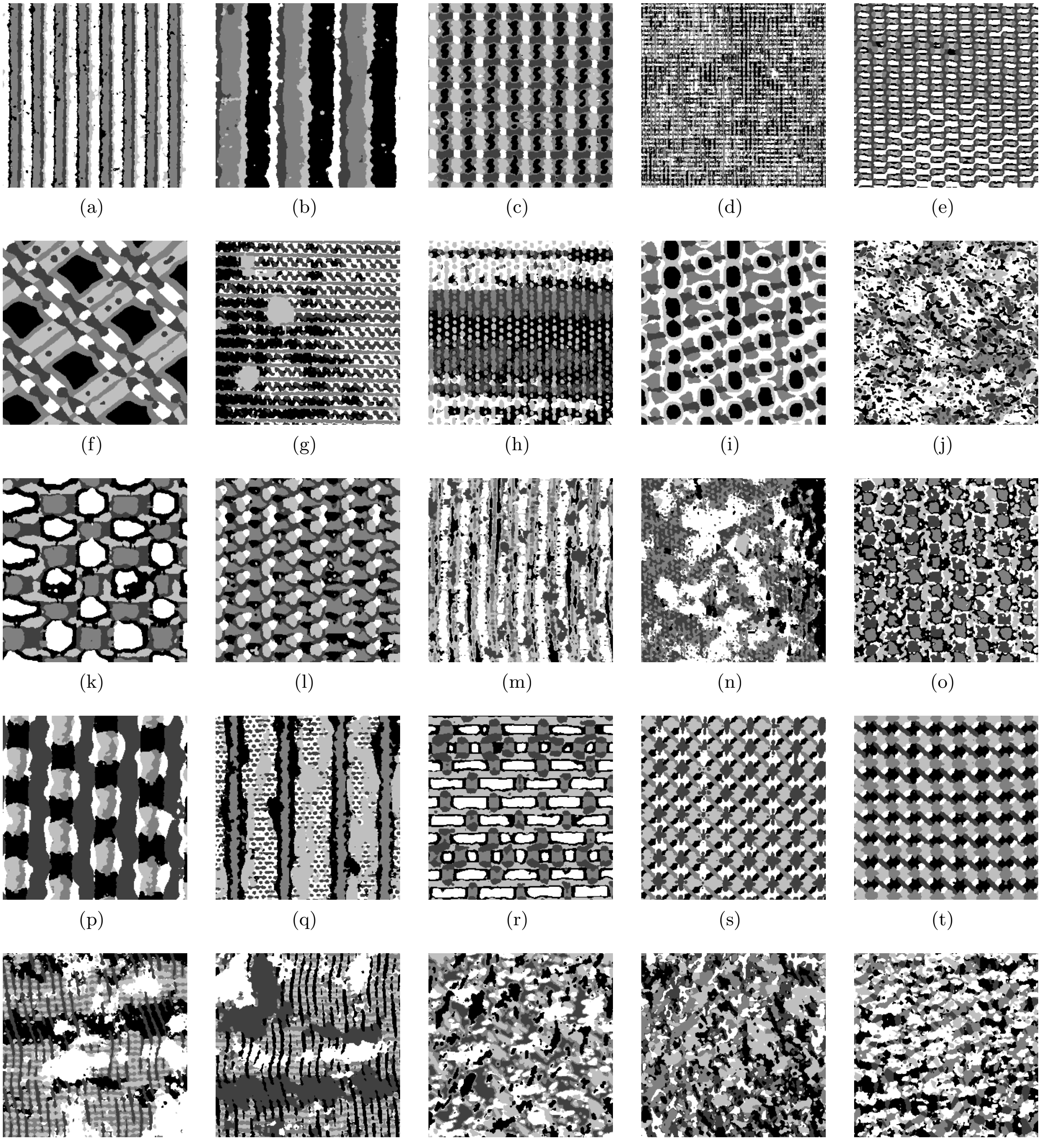


**Figure 6.** The motif assignments that result from using the single-level motif models to label the pixels in the texture images. These are the "images" that are used to learn the second-level models. (a) D016, (b) D017, (c) D020, (d) D021, (e) D034, (f) D047, (g) D053, (h) D055, (i) D056, (j) D057, (k) D064, (l) D065, (m) D076, (n) D077, (o) D082, (p) D083, (q) D085, (r) D095, (s) D101, (t) D102, (u) D103, (v) D014, (w) D109, (x) D110, and (y) D111. Only the center 256x256 pixels are shown for detail.